



the globus alliance  
www.globus.org

# Grids and the Globus Community

Dr. Jennifer M. Schopf

Argonne National Lab

<http://www.mcs.anl.gov/~jms/Talks/>



# What is a Grid?

- Resource sharing
  - Computers, storage, sensors, networks, ...
  - Sharing always conditional: issues of trust, policy, negotiation, payment, ...
- Coordinated problem solving
  - Beyond client-server: distributed data analysis, computation, collaboration, ...
- Dynamic, multi-institutional virtual orgs
  - Community overlays on classic org structures
  - Large or small, static or dynamic



# Why Is this Hard or Different?

- Lack of central control
  - Where things run
  - When they run
- Shared resources
  - Contention, variability
- Communication and coordination
  - Different sites implies different sys admins, users, institutional goals, and often socio-political constraints



## So Why Do It?

- Computations that need to be done with a time limit
- Data that can't fit on one site
- Data owned by multiple sites
- Applications that need to be run bigger, faster, more



# What Kinds of Applications?

- Computation intensive
  - Interactive simulation (climate modeling)
  - Large-scale simulation and analysis (galaxy formation, gravity waves, event simulation)
  - Engineering (parameter studies, linked models)
- Data intensive
  - Experimental data analysis (e.g., physics)
  - Image & sensor analysis (astronomy, climate)
- Distributed collaboration
  - Online instrumentation (microscopes, x-ray)
  - Remote visualization (climate studies, biology)
  - Engineering (large-scale structural testing)



## Globus is...

- A collection of solutions to problems that come up frequently when building collaborative distributed applications
- Software for Grid infrastructure
  - Service enable new & existing resources
  - Uniform abstractions & mechanisms
- Tools to build applications that exploit Grid infrastructure
  - Registries, security, data management, ...
- Open source & open standards
  - Each empowers the other
- Enabler of a rich tool & service ecosystem



# Globus is an Hour Glass

- Local sites have their own policies, installs – heterogeneity!

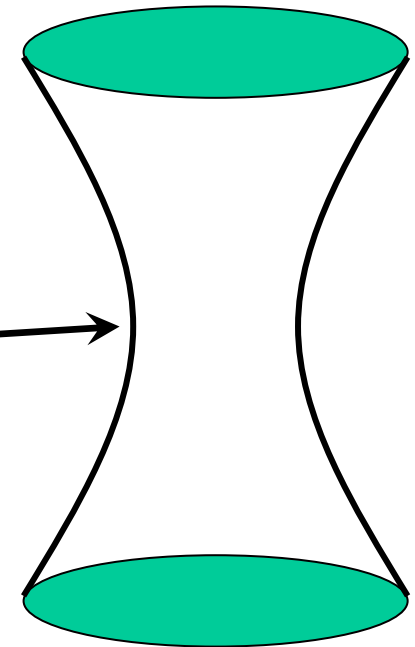
- Queuing systems, monitors, network protocols, etc

- Globus unifies – standards!

- Build on Web services
- Use WS-RF, WS-Notification to represent/access state
- Common management abstractions & interfaces

Higher-Level Services  
and Users

Standard  
Interfaces



Local heterogeneity



# Globus is a Building Block

- Basic components for Grid functionality
  - Not turnkey solutions, but building blocks & tools for application developers & system integrators
- Highest-level services are often application specific, we let aps concentrate there
- Easier to reuse than to reinvent
  - Compatibility with other Grid systems comes for free
- We provide basic infrastructure to get you one step closer





- Governance model based on Apache Jakarta
  - Consensus based decision making
- Globus software is organized as several dozen “Globus Projects”
  - Each project has its own “Committers” responsible for their products
  - Cross-project coordination through shared interactions and committers meetings
- A “Globus Management Committee”
  - Overall guidance and conflict resolution

- [Welcome](#)
- [List of projects](#)
- [Guidelines](#)
- [Infrastructure](#)
- [How to contribute](#)
- [GlobDev events](#)
- [Recent changes](#)
- [GlobDev FAQ](#)

#### *common runtime projects*

- [C Core Utilities](#)
- [C WS Core](#)
- [CoG jglobus](#)
- [Core WS Schema](#)
- [Java WS Core](#)
- [Python Core](#)
- [XIO](#)

#### *data projects*

- [GridFTP](#)
- [OGSA-DAI](#)
- [Reliable File Transfer](#)
- [Replica Location](#)

#### *execution projects*

- [GRAM](#)

#### *information projects*

- [MDS4](#)

#### *security projects*

- [C Security](#)
- [CAS/SAML Utilities](#)
- [Delegation Service](#)

## Welcome

This is the new home Globus software development; it is still under construction. The current status of our efforts to build this environment can be found [on this page](#). Comments regarding this site can be sent to [info@globus.org](mailto:info@globus.org). Thank you for your interest in Globus development!

Globus was first established as an open source software project in 1996. Since that time, the Globus development team has expanded from a few individuals to a distributed, international community. In response to this growth, the Globus community (the "Globus Alliance") established in October 2005 a new source code development *infrastructure* and meritocratic *governance model*, which together make the process by which a developer joins the Globus community both easier and more transparent.

The Globus governance model and infrastructure are based on those of [Apache Jakarta](#). In brief, the governance model places control over each individual software component ([project](#)) in the hands of its most active and respected [contributors](#) (*committers*), with a [Globus Management Committee](#) (GMC) providing overall guidance and conflict resolution. The infrastructure comprises [repositories](#), [email lists](#), Wikis, and [bug trackers](#) configured to support per-project community access and management.

For more information, see:

- The [Globus Alliance Guidelines](#), which address various aspects of the Globus governance model and the Globus community.
- A description of the Globus Alliance [Infrastructure](#).
- A list of current Globus projects.
- Information about Globus community events.
- The [conventions and guidelines](#) that apply to contributions

Guidelines  
(Apache  
Jakarta)

Infrastructure  
(CVS, email,  
bugzilla, Wiki)

Projects  
Include

...



# Globus Software: [dev.globus.org](http://dev.globus.org)

## Globus Projects

MPICH G2

GridWay

Incubation  
Mgmt

Java  
Runtime

C  
Runtime

Python  
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-  
OpenSSH

GRAM

OGSA-DAI

Data  
Rep

GridFTP

Reliable  
File  
Transfer

GT4

Replica  
Location

MDS4

GT4 Docs

## Incubator Projects

GDTE

GridShib

OGRO

Introduce

PURSE

HOC-SA

Swift

MEDICUS

UGP

LRMA

GEMICA

Cog WF

Dyn Acct

WEEP

Virt WkSp

Gavia JSC

Gavia MS

DDM

SGGC

MonMan

NetLogger

Metrics

ServMark

Common  
Runtime

Security

Execution  
Mgmt

Data Mgmt

Info  
Services

Other



# Globus Technology Areas

- Core runtime
  - Infrastructure for building new services
- Security
  - Apply uniform policy across distinct systems
- Execution management
  - Provision, deploy, & manage services
- Data management
  - Discover, transfer, & access large data
- Monitoring
  - Discover & monitor dynamic services

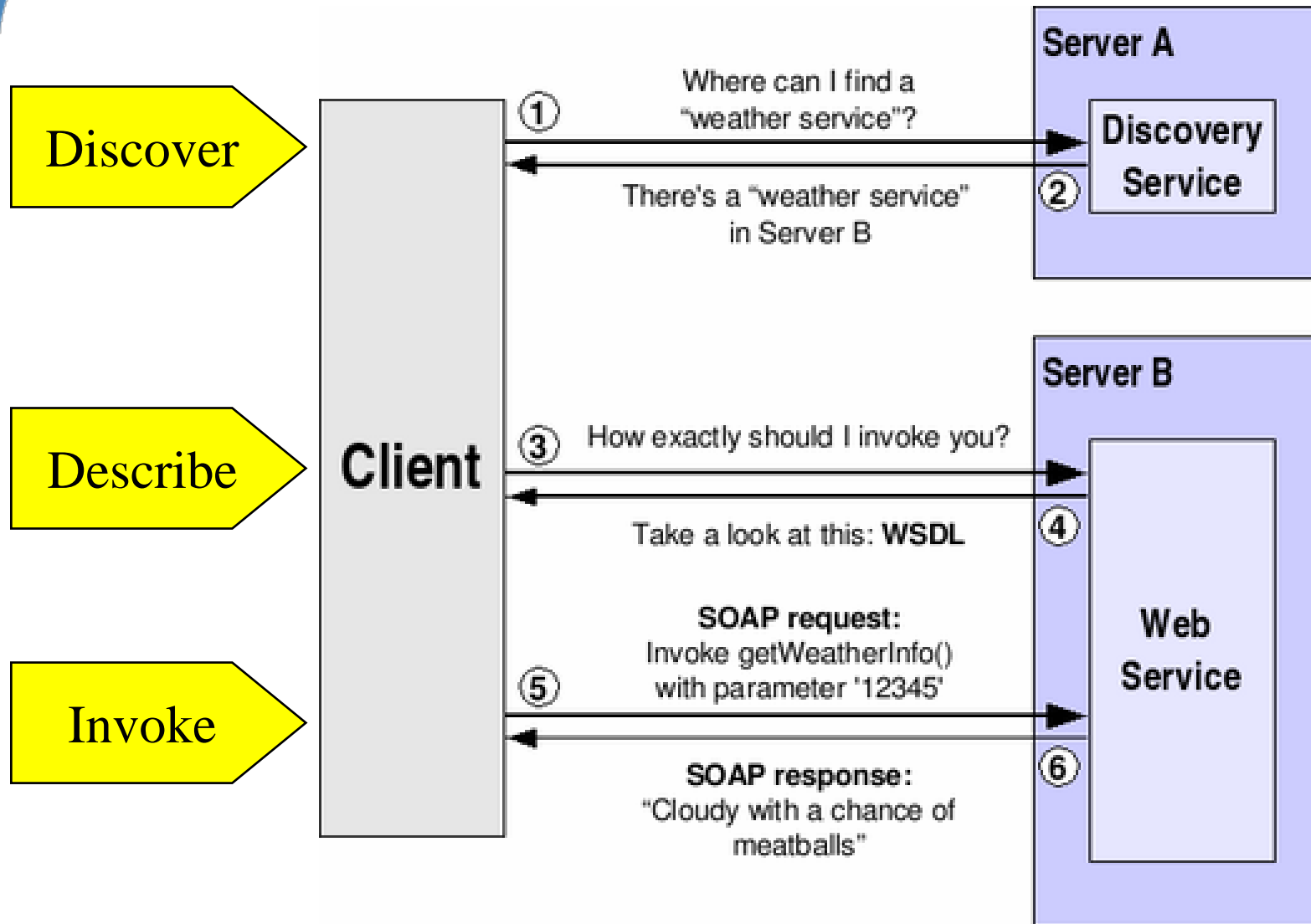


## Core Runtime Provides Web Service Basics

- Web services are platform independent and language independent
  - Client and server program can be written in diff langs, run in diff envt's and still interact
- Web services describe themselves
  - Once located you can ask it how to use it
- Web service is *\*not\** a website
  - Web service is accessed by sw, not humans
- Web services are ideal for loosely coupled systems
  - Unlike CORBA, EJB, etc.



# Real Web Service Invocation





## Need For Standards

- Web services are self describing using WSDL
- But we'd really like is a common way to
  - Name and do bindings
  - Start and end services
  - Query, subscription, and notification
  - Share error messages



# WSRF & WS-Notification

- Naming and bindings (basis for virtualization)
  - Every resource can be uniquely referenced, and has one or more associated services for interacting with it
- Lifecycle (basis for fault resilient state management)
  - Resources created by services following factory pattern
  - Resources destroyed immediately or scheduled
- Information model (basis for monitoring & discovery)
  - Resource properties associated with resources
  - Operations for querying and setting this info
  - Asynchronous notification of changes to properties
- Service Groups (basis for registries & collective svcs)
  - Group membership rules & membership management
- Base Fault type



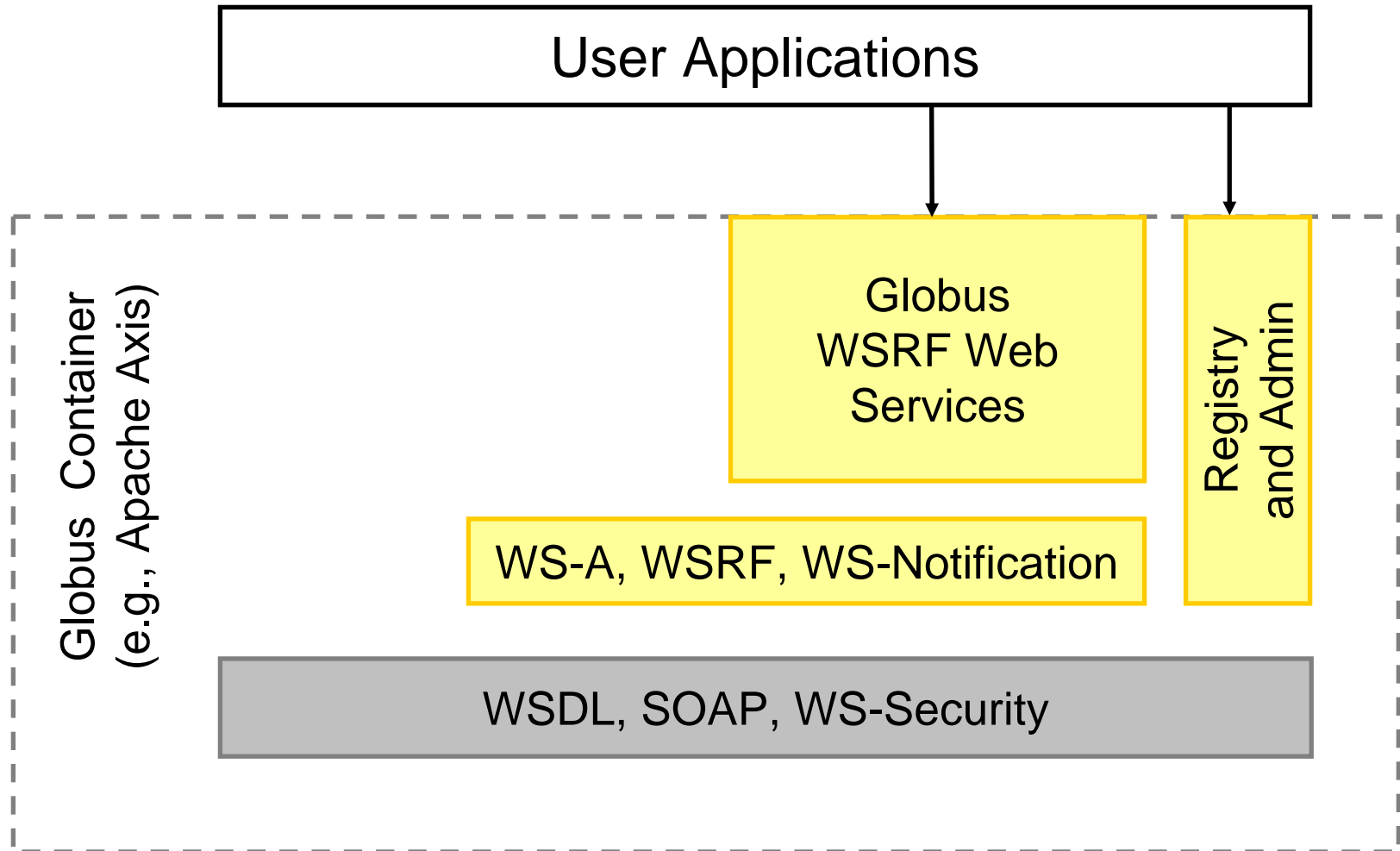


## WSRF vs XML/SOAP

- The definition of WSRF means that the Grid and Web services communities can move forward on a common base
- Why Not Just Use XML/SOAP?
  - WSRF and WS-N *are* just XML and SOAP
  - WSRF and WS-N are just Web services
- Benefits of following the specs:
  - These patterns represent best practices that have been learned in many Grid applications
  - There is a community behind them
  - Why reinvent the wheel?
  - Standards facilitate interoperability

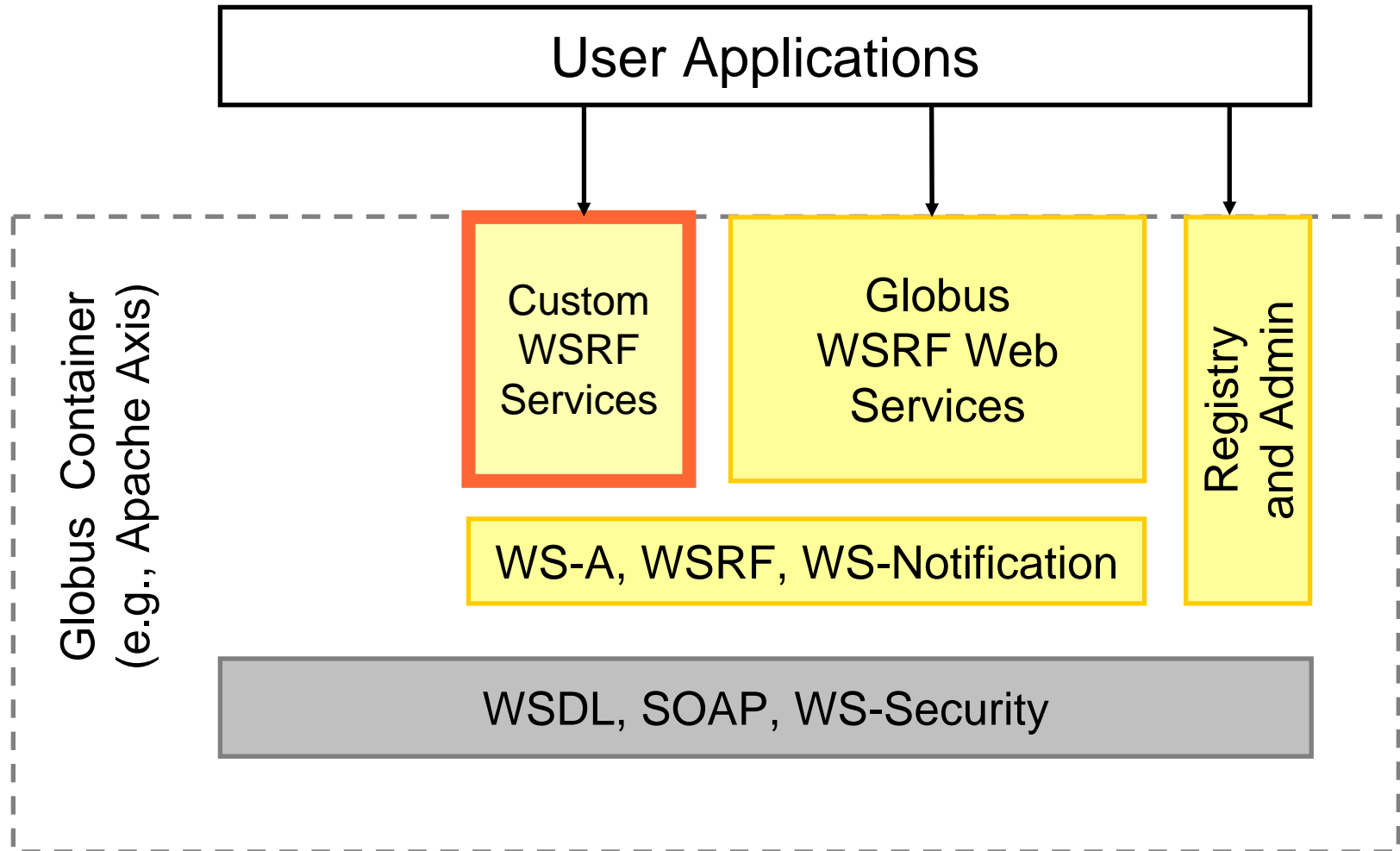


# Globus and Web Services



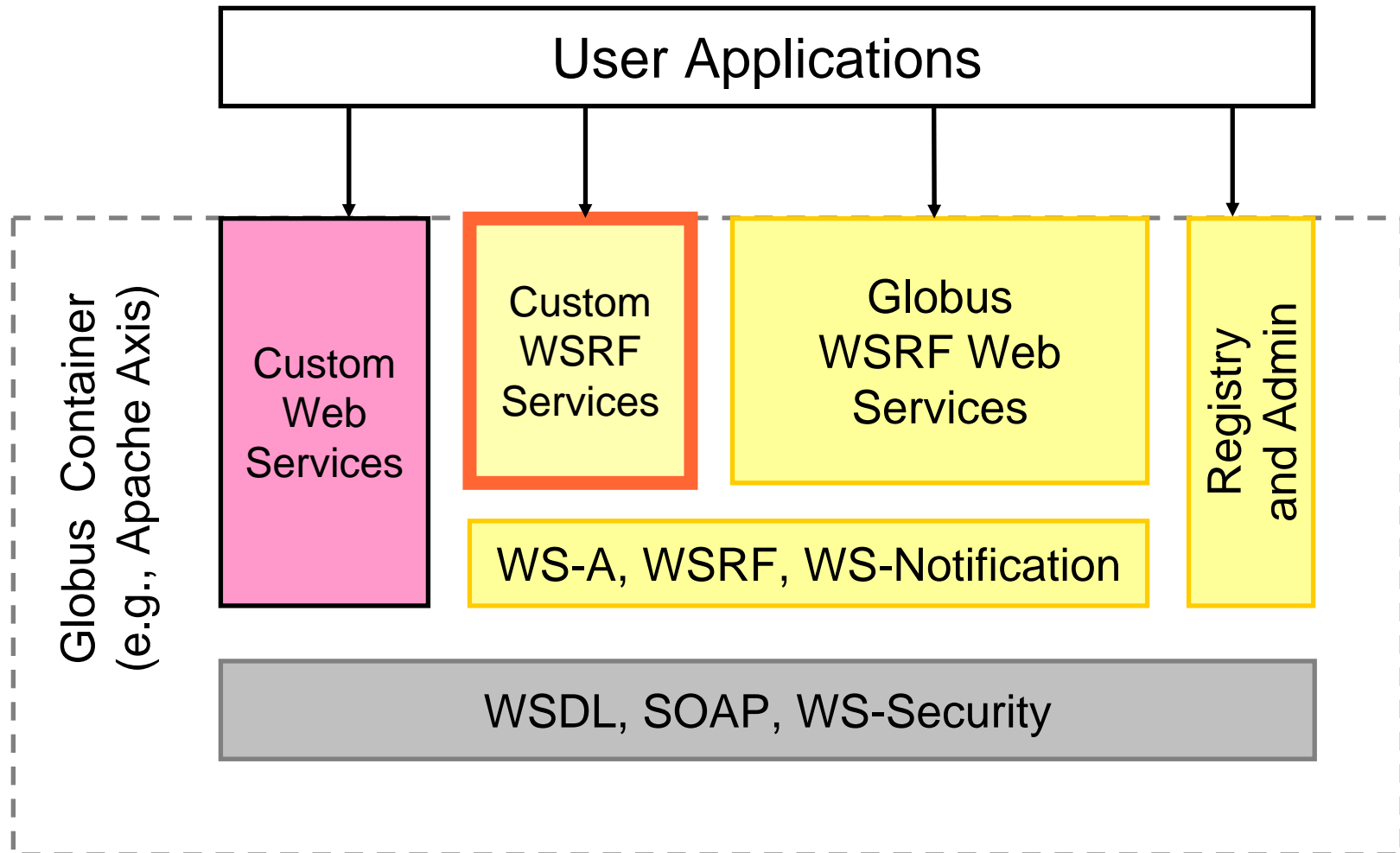


# Globus and Web Services





# Globus and Web Services





# Globus Technology Areas

- Core runtime
  - Infrastructure for building new services
- Security
  - Apply uniform policy across distinct systems
- Execution management
  - Provision, deploy, & manage services
- Data management
  - Discover, transfer, & access large data
- Monitoring
  - Discover & monitor dynamic services



# Grid Security Concerns

- Control access to shared services
  - Address autonomous management, e.g., different policy in different work groups
- Support multi-user collaborations
  - Federate through mutually trusted services
  - Local policy authorities rule
- Allow users and application communities to set up dynamic trust domains
  - Personal/VO collection of resources working together based on trust of user/VO



# Globus Security Tools

- Basic Grid Security Mechanisms
- Certificate Generation Tools
- Certificate Management Tools
  - Getting users “registered” to use a Grid
  - Getting Grid credentials to wherever they’re needed in the system
- Authorization/Access Control Tools
  - Storing and providing access to system-wide authorization information

## Execution Management: GRAM

- GRAM: Grid Resource Allocation Manager
- A uniform service interface for remote job submission and control
  - Unix, Condor, LSF, PBS, SGE, ...
- More generally: interface for process execution management
  - Lay down execution environment
  - Stage data
  - Monitor & manage lifecycle
  - Kill it, clean up



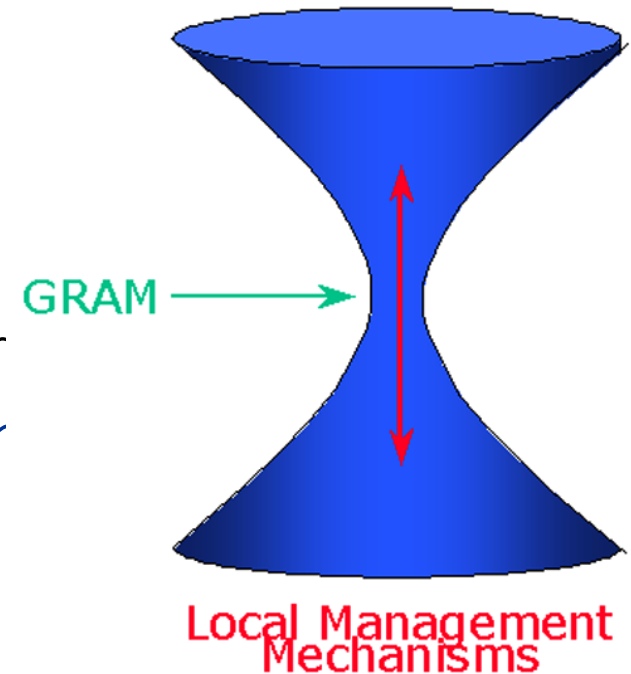


## GRAM4 (aka WS GRAM)

- 2nd-generation WS implementation optimized for performance, flexibility, stability, scalability
- Streamlined critical path
  - Use only what you need
- Flexible credential management
  - Credential cache & delegation ser
- GridFTP & RFT used for data oper
  - Data staging & streaming output
  - Eliminates redundant GASS code
- GRAM is *not* a scheduler.
  - Used as a front-end to schedulers,

### Applications

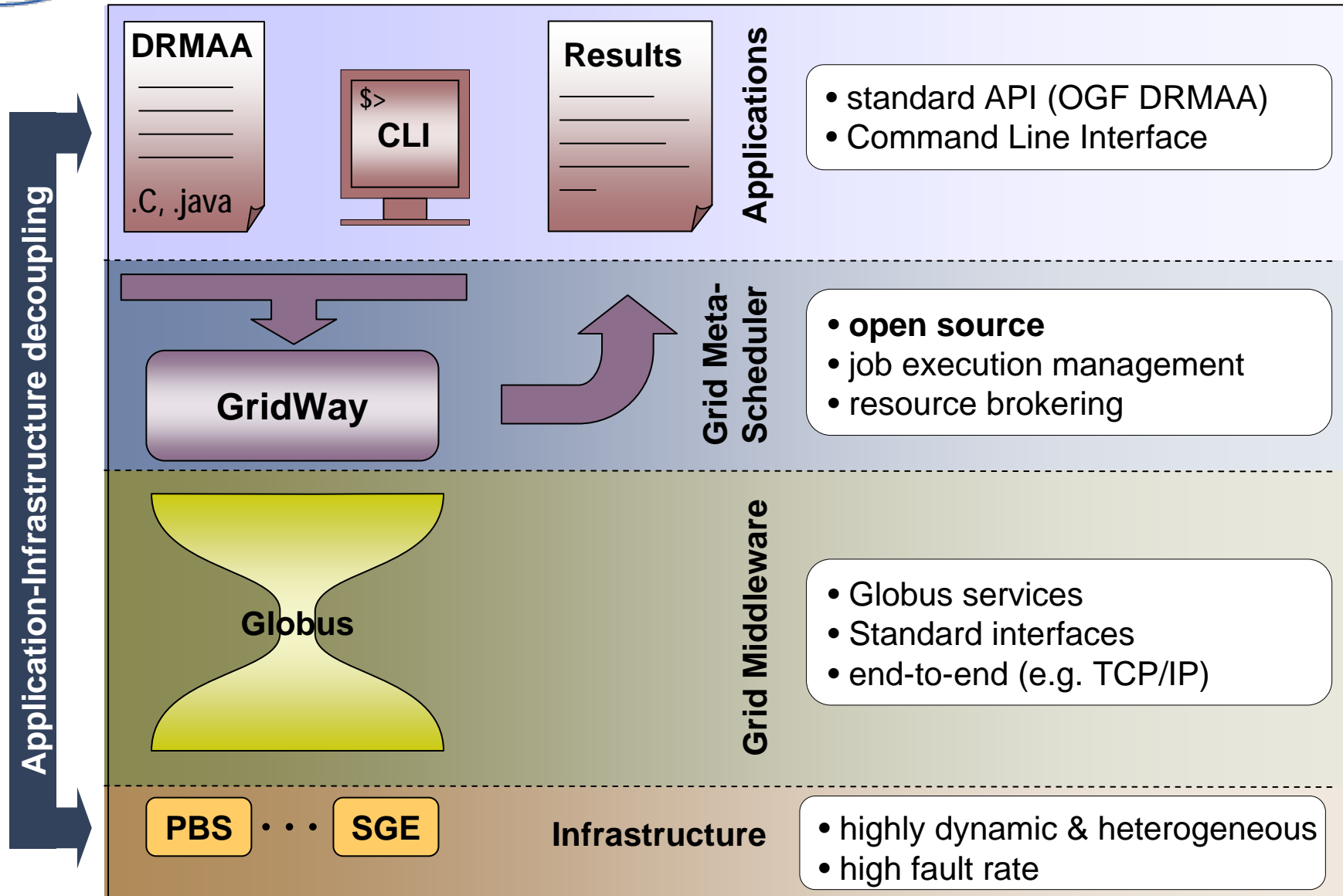
Metaschedulers, Brokers





# GridWay Meta-Scheduler

- Scheduler virtualization layer on top of Globus services
  - A LRM-like environment for submitting, monitoring, and controlling jobs
  - A way to submit jobs to the Grid, without having to worry about the details of exactly which local resource will run the job
  - A policy-driven job scheduler, implementing a variety of access and Grid-aware load balancing policies
  - Accounting





# GT4 Data Management

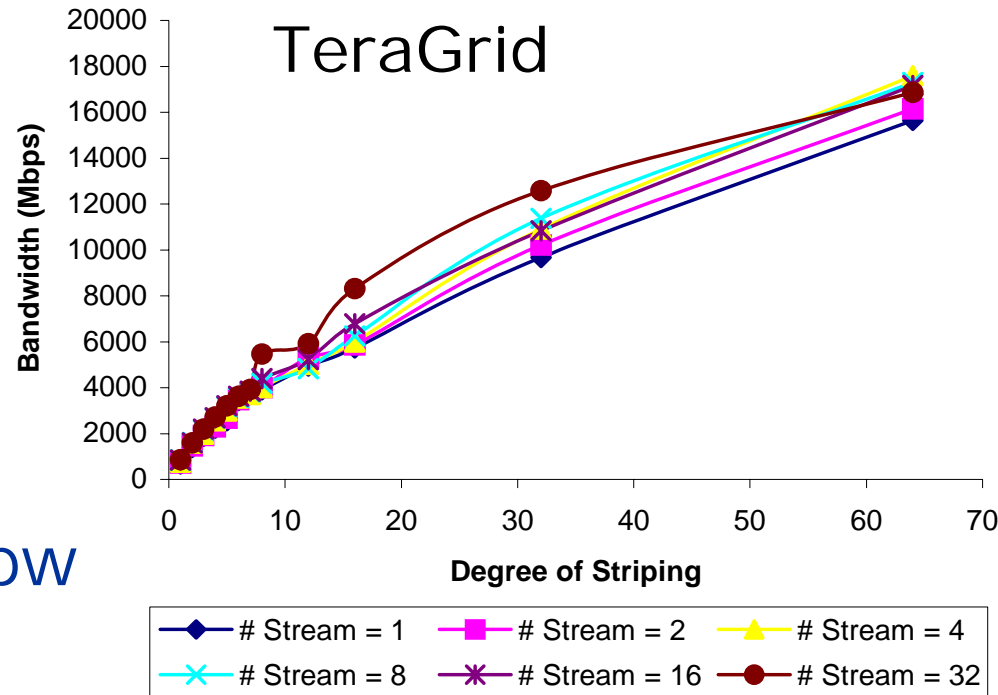
- **Stage/move** large data to/from nodes
  - GridFTP, Reliable File Transfer (RFT)
  - Alone, and integrated with GRAM
- **Locate** data of interest
  - Replica Location Service (RLS)
- **Replicate** data for performance/reliability
  - Distributed Replication Service (DRS)
- Provide **access** to diverse data sources
  - File systems, parallel file systems, hierarchical storage: GridFTP
  - Databases: OGSA DAI



## GridFTP in GT4

- A high-performance, secure, reliable data transfer protocol optimized for high-bw wide-area networks
- GSI support for security
- 3<sup>rd</sup> party and partial file transfer support
- IPv6 Support
- XIO for different transports
- Parallelism and striping → multi-Gb/sec wide area transport

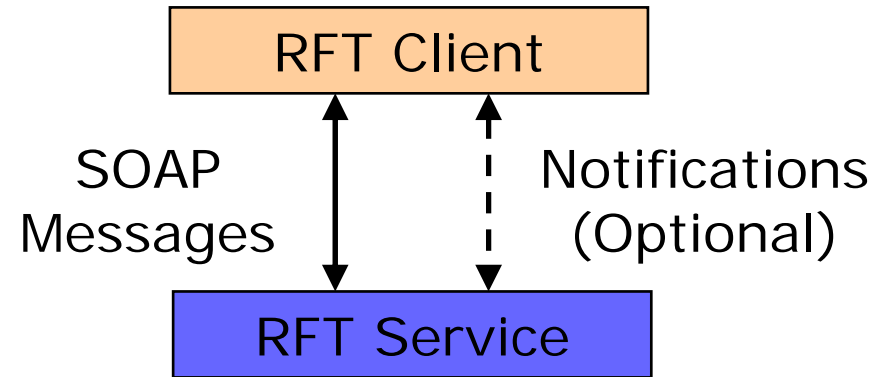
### Disk-to-disk on TeraGrid





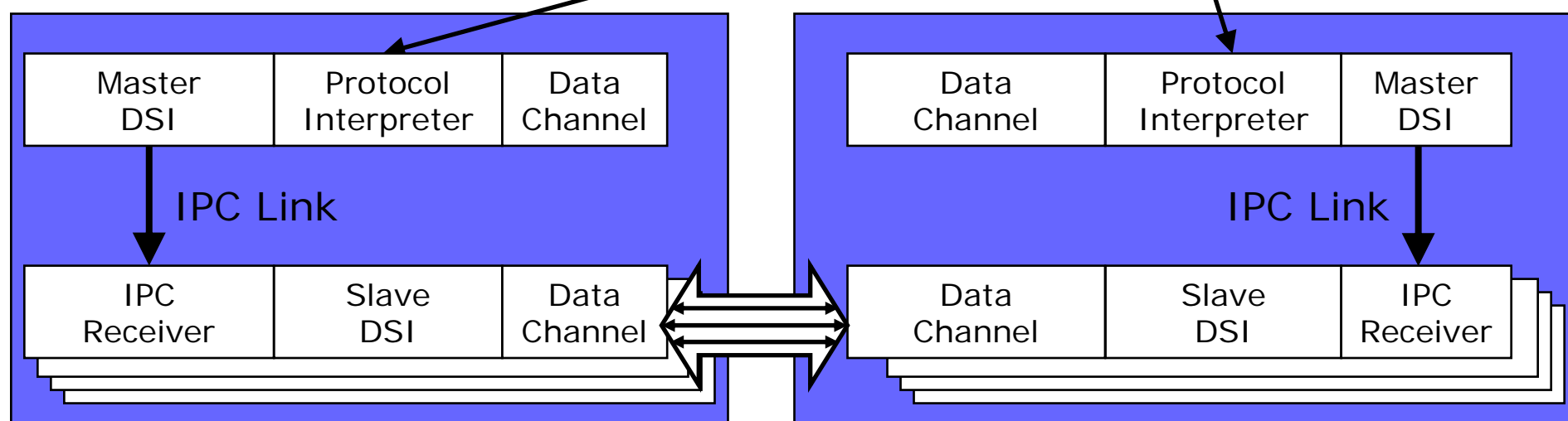
# Reliable File Transfer

- Fire-and-forget transfer
- Web services interface
- Many files & directories
- Integrated failure recovery
- Has transferred 900K files



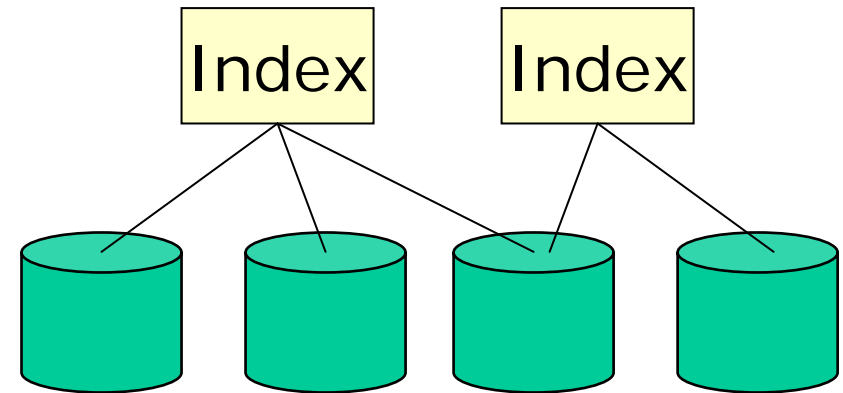
GridFTP Server

GridFTP Server



# Replica Location Service

- Identify location of files via logical to physical name map
- Distributed indexing of names, fault tolerant update protocols
- New WS-RF version available
- Managing ~40 million files across ~10 sites



Local DB	Update send (secs)	Bloom filter (secs)	Bloom filter (bits)
10K	<1	2	1 M
1 M	2	24	10 M
5 M	7	175	50 M



- Grid Interfaces to Databases
  - Data access
    - > Relational & XML Databases, semi-structured files
  - Data integration
    - > Multiple data delivery mechanisms, data translation
- Extensible & Efficient framework
  - Request documents contain multiple tasks
    - > A task = execution of an activity
    - > Group work to enable efficient operation
  - Extensible set of activities
    - > > 30 predefined, framework for writing your own
  - Moves computation to data
  - Pipelined and streaming evaluation
  - Concurrent task evaluation





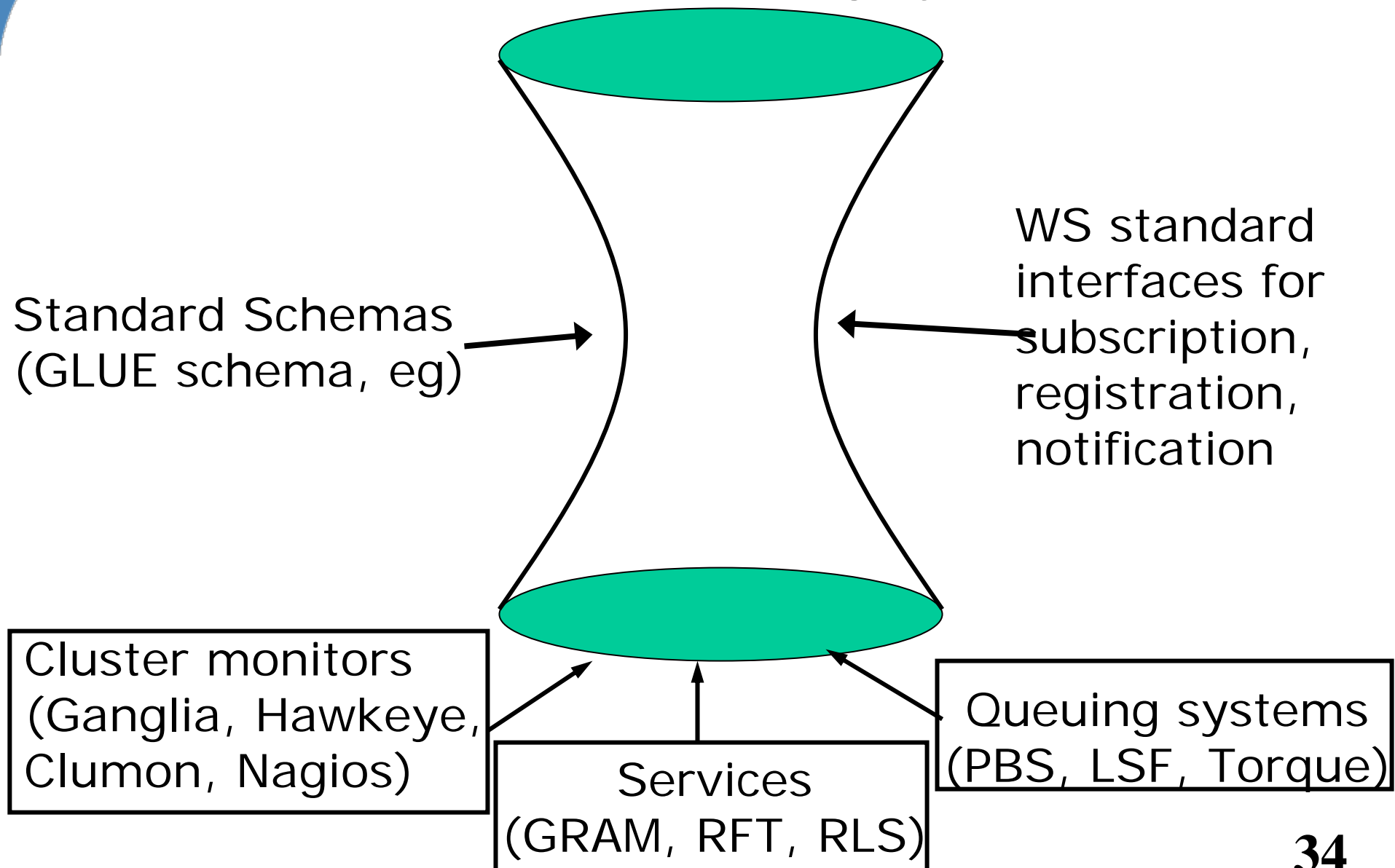
# Monitoring and Discovery System (MDS4)

- Grid-level monitoring system
  - Aid user/agent to identify host(s) on which to run an application
  - Warn on errors
- Uses standard interfaces to provide publishing of data, discovery, and data access, including subscription/notification
  - WS-ResourceProperties, WS-BaseNotification, WS-ServiceGroup
- Functions as an hourglass to provide a common interface to lower-level monitoring tools



## Information Users :

Schedulers, Portals, Warning Systems, etc.





# Globus Technology Areas

- Core runtime
  - Infrastructure for building new services
- Security
  - Apply uniform policy across distinct systems
- Execution management
  - Provision, deploy, & manage services
- Data management
  - Discover, transfer, & access large data
- Monitoring
  - Discover & monitor dynamic services



# Non-Technology Projects

- Incubation Projects
  - Incubation management project
  - And any new projects wanting to join
- Distribution Projects
  - Globus Toolkit Distribution
- Documentation Projects
  - GT Release Manuals



# Globus Software: [dev.globus.org](http://dev.globus.org)

## Globus Projects

MPICH G2

GridWay

Incubation  
Mgmt

Java  
Runtime

C  
Runtime

Python  
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-  
OpenSSH

GRAM

OGSA-DAI

Data  
Rep

GridFTP

Reliable  
File  
Transfer

GT4

Replica  
Location

MDS4

GT4 Docs

## Incubator Projects

Swift

GEMICA

MonMan

GAARDS

MEDICUS

Cog WF

Virt WkSp

NetLogger

GDTE

GridShib

OGRO

UGP

Dyn Acct

Gavia JSC

DDM

Metrics

Introduce

PURSE

HOC-SA

LRMA

WEEP

Gavia MS

SGGC

ServMark

**Common  
Runtime**

**Security**

**Execution  
Mgmt**

**Data Mgmt**

**Info  
Services**

**Other**

## Incubator Process in dev.globus

- Entry point for new Globus projects
- Incubator Management Project (IMP)
  - Oversees incubator process from first contact to becoming a Globus project
  - Quarterly reviews of current projects

[http://dev.globus.org/wiki/Incubator/Incubator\\_Process](http://dev.globus.org/wiki/Incubator/Incubator_Process)



## 24 Active Incubator Projects

- CoG Workflow
- Distributed Data Management (DDM)
- Dynamic Accounts
- Grid Authentication and Authorization with Reliably Distributed Services (GAARDS)
- Gavia-Meta Scheduler
- Gavia- Job Submission Client
- Grid Development Tools for Eclipse (GDTE)
- Grid Execution Mgmt. for Legacy Code Apps. (GEMLCA)
- GridShib
- Higher Order Component Service Architecture (HOC-SA)
- Introduce
- Local Resource Manager Adaptors (LRMA)
- MEDICUS (Medical Imaging and Computing for Unified Information Sharing)
- Metrics
- MonMan
- NetLogger
- Open GRid OCSP (Online Certificate Status Protocol)
- Portal-based User Registration Service (PURSe)
- ServMark
- SJTU GridFTP GUI Client (SGGC)
- Swift
- UCLA Grid Portal Software (UGP)
- Workflow Enactment Engine Project (WEEP)
- Virtual Workspaces

## Active Committers from 28 Institutions

- Aachen Univ. (Germany)
- Argonne National Laboratory
- CANARIE (Canada)
- CertiVeR
- Children's Hospital Los Angeles
- Delft Univ. (The Netherlands)
- Indiana Univ.
- Kungl. Tekniska Högskolan (Sweden)
- Lawrence Berkeley National Lab
- Leibniz Supercomputing Center (Germany)
- NCSA
- National Research Council of Canada
- Ohio State Univ.
- Semantic Bits
- Shanghai Jiao Tong University (China)
- Univ. of British Columbia (Canada)
- UCLA
- Univ. of Chicago
- Univ. of Delaware
- Univ. of Marburg (Germany)
- Univ. of Muenster (Germany)
- Univ. Politecnica de Catalunya (Spain)
- Univ. of Rochester
- USC Information Sciences Institute
- Univ. of Victoria (Canada)
- Univ. of Vienna (Austria)
- Univ. of Westminster (UK)
- Univa Corp.





## Globus Projects

MPICH G2

GridWay

Incubation  
Mgmt

Java  
Runtime

C  
Runtime

Python  
Runtime

Delegation

CAS

C Sec

MyProxy

GSI-  
OpenSSH

GRAM

OGSA-DAI

Data  
Rep

GridFTP

Reliable  
File  
Transfer

GT4

Replica  
Location

MDS4

GT4 Docs

## Incubator Projects

Incubator Projects				Swift	GEMICA				MonMan
	GAARDS			MEDICUS	Cog WF	Virt WkSp			NetLogger
	GDTE	GridShib	OGRO	UGP	Dyn Acct	Gavia JSC	DDM	Metrics	
	Introduce	PURSE	HOC-SA	LRMA	WEEP	Gavia MS	SGGC	ServMark	

Common  
Runtime

Security

Execution  
Mgmt

Data Mgmt

Info  
Services

Other



## Globus Projects

MPICH G2

GridWay

Incubation  
Mgmt

Java  
Runtime

Delegation

MyProxy

OGSA-DAI

GT4

C  
Runtime

CAS

GSI-  
OpenSSH

Data  
Rep

Replica  
Location

Python  
Runtime

C Sec

GRAM

GridFTP

MDS4

Reliable  
File  
Transfer

GT4 Docs

## Incubator Projects

Swift

GEMLCA

MonMan

GAARDS

MEDICUS

Cog WF

Virt WkSp

NetLogger

GDTE

GridShib

OGRO

UGP

Dyn Acct

Gavia JSC

DDM

Metrics

Introduce

PURSE

HOC-SA

LRMA

WEEP

Gavia MS

SGGC

ServMark

Common  
Runtime

Security

Execution  
Mgmt

Data Mgmt

Info  
Services

Other



## GT4 Distribution

- Usability, reliability
  - All components meet a quality standard
  - Testing, logging, coding standards
  - Documentation at acceptable quality level
  - Guarantee that interfaces won't change within a major version (4.0.1 == 4.0.any)
- Consistency with latest standards (WS-\*, WSRF, WS-N, etc.) and Apache platform
  - WS-I Basic Profile compliant
  - WS-I Basic Security Profile compliant



# Globus User Community

- Large & diverse
  - 10s of national Grids, 100s of applications, 1000s of users; probably much more
  - Every continent except Antarctica
  - Applications ranging across many fields
  - Dozens (at least) of commercial deployments
- Successful
  - Many production systems doing real work
  - Many applications producing real results
  - Hundreds of papers published because of grid deployments
- Smart, energetic, demanding
  - Constant stream of new use cases & tools



## How Can You Contribute? Create a New Project

- Do you have a project you'd like to contribute?
- Does your software solve a problem you think the Globus community would be interested in?
- Contact [incubator-committers@globus.org](mailto:incubator-committers@globus.org)



## How Can You Contribute? Help an Existing Project

- Contribute code, documentation, design ideas, and feature requests
- Joining the mailing lists
  - \*-dev, \*-user, \*-commit for each project
  - See the project wiki page at [dev.globus.org](http://dev.globus.org)
- Chime in at any time
- Regular contributors can become committers, with a role in defining project directions

[http://dev.globus.org/wiki/How\\_to\\_contribute](http://dev.globus.org/wiki/How_to_contribute)



## Globus Next Steps

- Expanded open source Grid infrastructure
  - Updates for current standards
  - New services for data management, security, VO management, troubleshooting
  - End-user tools for application development
  - Virtualization
- Some infrastructure work
  - Outside projects joining Globus
  - Expanded outreach: [outreach@globus.org](mailto:outreach@globus.org)
- And of course responding to user requests for other short-term needs



## For More Information

- Jennifer Schopf
  - [jms@mcs.anl.gov](mailto:jms@mcs.anl.gov)
  - <http://www.mcs.anl.gov/~jms>
- Globus Alliance
  - <http://www.globus.org>
- Dev.globus
  - <http://dev.globus.org>
- Upcoming Events
  - <http://dev.globus.org/wiki/Outreach>